

JEWELS IN JUNK DNA

Using cross-species sequence comparisons to gain insights into the functions of gene deserts.

• **Marcelo A. Nobrega** - manobrega@lbl.gov

• **Ivan V. Ovcharenko** - ivovcharenko@lbl.gov

Genome Sciences Department
Lawrence Berkeley National Laboratory

-
- 2-5% of DNA is translated
 - 25-30% of DNA is transcribed
 - 60-70% of the genome contains non-genic DNA

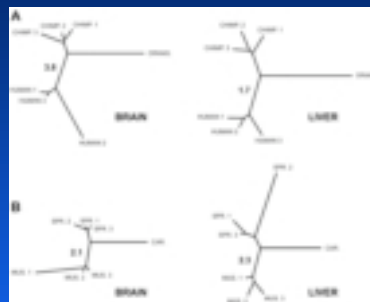
Why do we have a 2,900,000,000 bp Genome?

C-value Paradox

Genome size does not correlate with organismal complexity



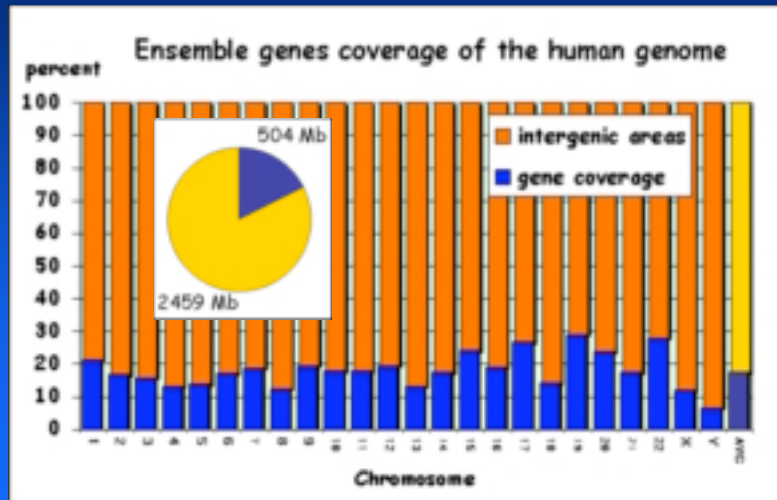
Regulation of gene expression governs speciation



Differences

	Genes	Qualitative	Quantitative
Human-Chimpanzee	538	41(7.6%)	169(31.4%)

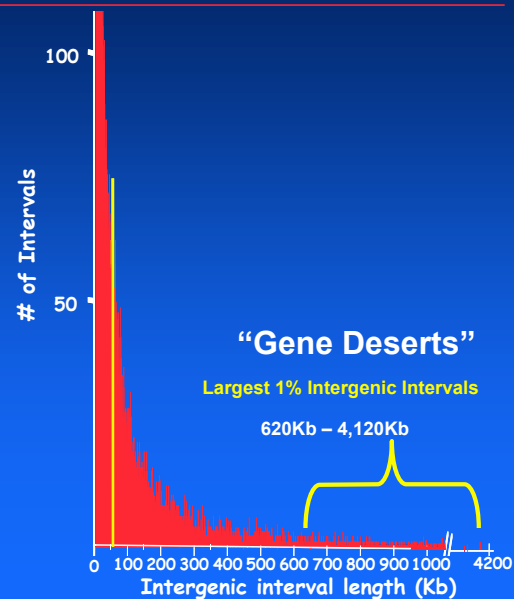
Gene coverage of the human genome



Defining “Gene Deserts”

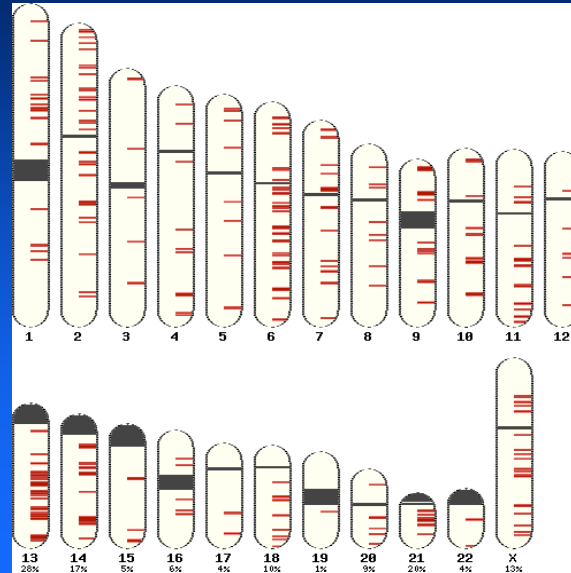
- 1 – Calculate Intergenic Lengths
EMSEMBL- 21,978 genes
REFSEQ Annotation- 12,439 genes

- 2 – Exclude Heterochromatic DNA
& Clone Gaps.

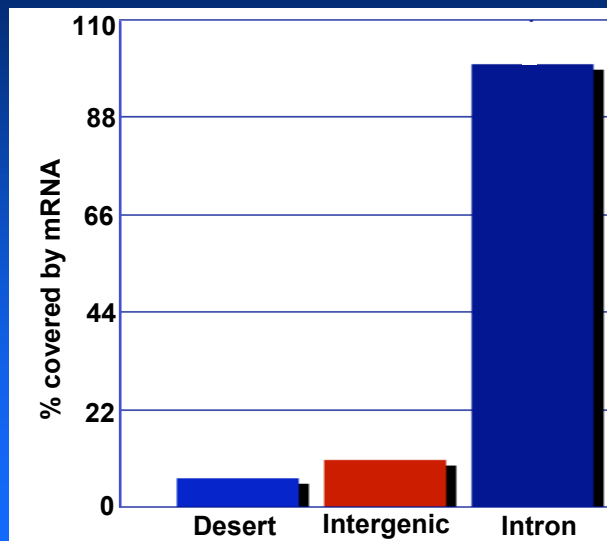


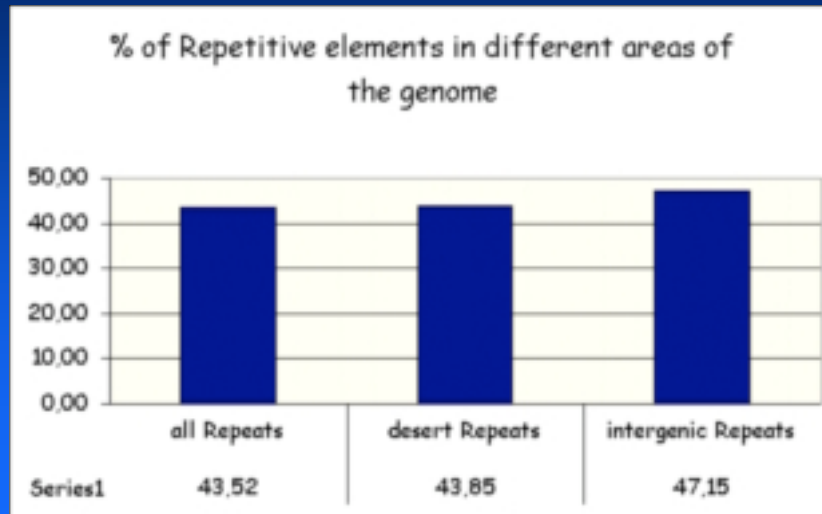
Distribution of Human “Gene Deserts”

- Total #
• 234 deserts
- Size Range
• 680 – 4,120 Kbp
- % of Genome
• 9% (277 Mbp)

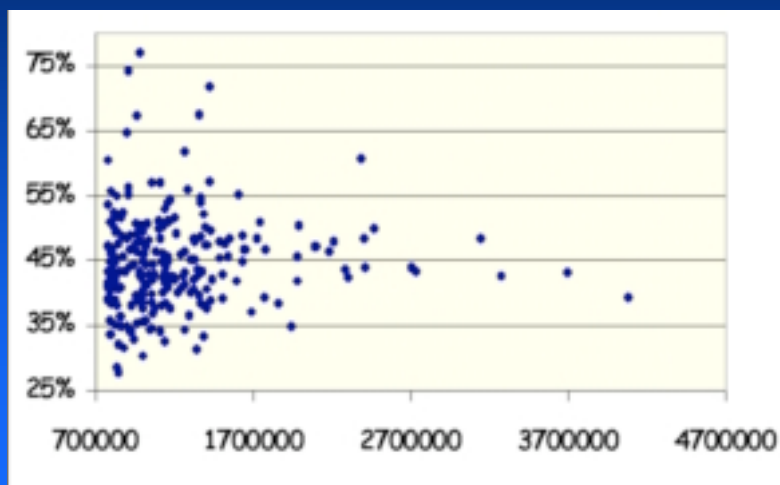


mRNA Content of Genomic Regions

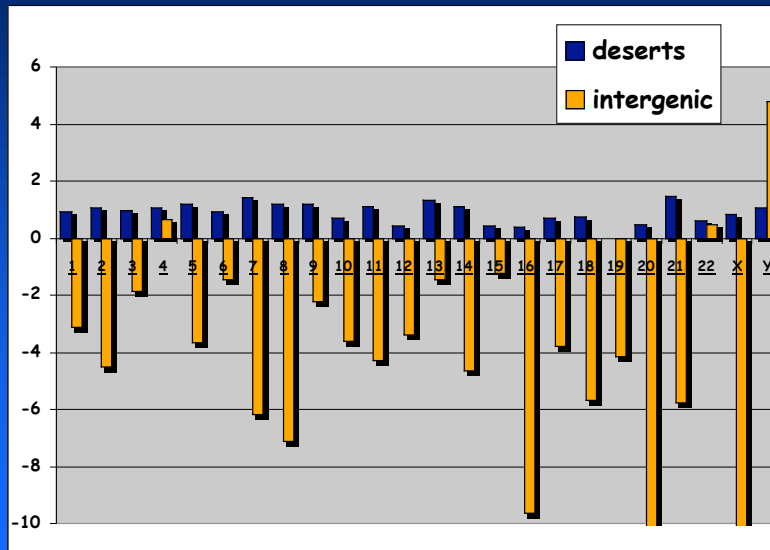




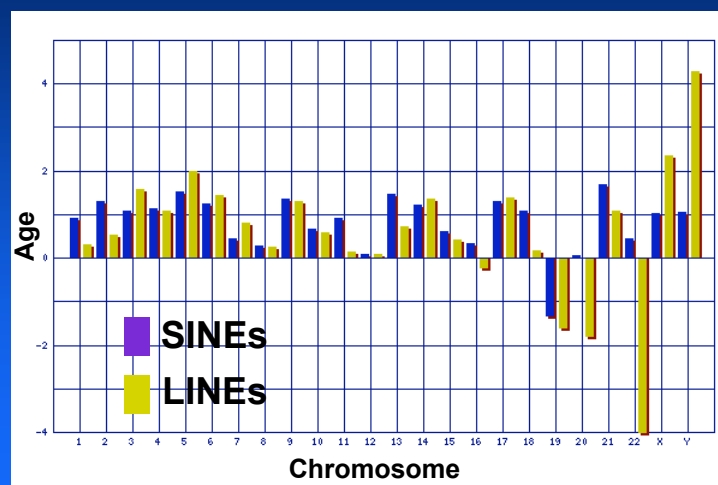
Correlation between repeat content and desert length



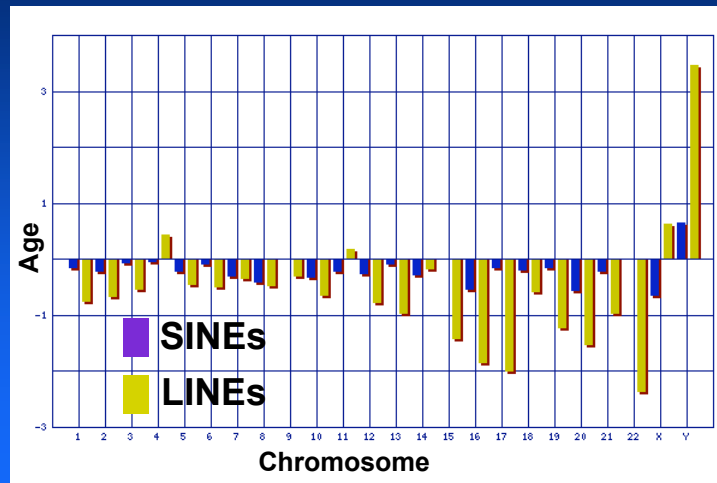
Age of SINES in Intergenic DNA



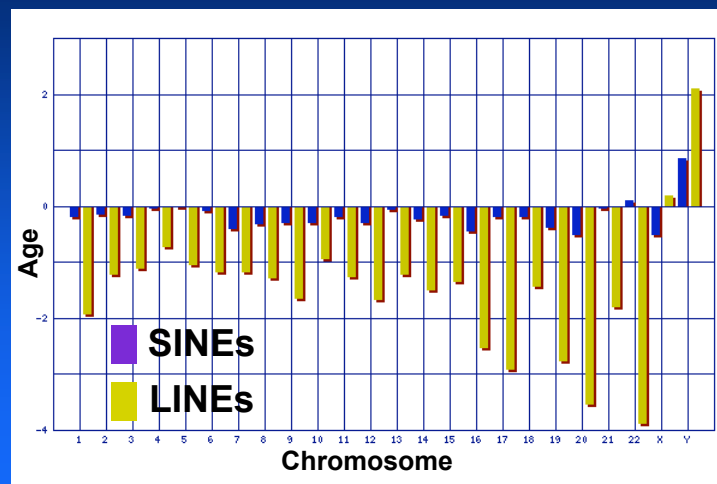
Age of Repetitive Elements in Gene Deserts



Age of Repetitive Elements in Intergenic Regions



Age of Repetitive Elements in Introns



Summary of Human Gene Deserts Findings

- 1- Don't appear to result from the expansion of genomic segments due to transposon insertions
- 2 – Contain primarily younger repetitive elements than the intergenic or intronic fraction of the genome
- 3 – The amount of younger insertions doesn't predict length of the gene desert

Can the younger age of insertions reflect a higher degree of “housekeeping” deletions in deserts?

Are Human “Gene Deserts” Also Deserts in Mouse?

Strategy for Identifying Mouse Gene Deserts

1 –BLAT Comparison of Human “Gene Deserts” to Mouse Genome Assembly

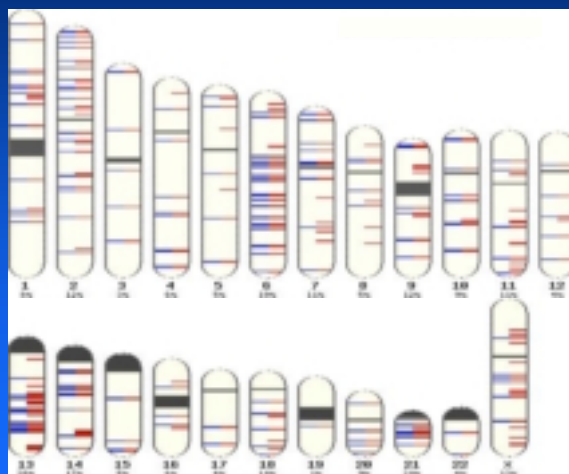
2 – Search for predicted genes in orthologous mouse DNA:

- Public Mouse Assembly
RefSeq Annotation(8,285 genes)
- Celera Mouse Assembly
Less than 2 lines of evidence for gene prediction

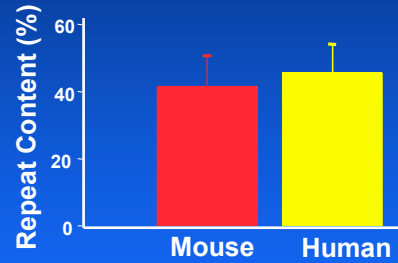
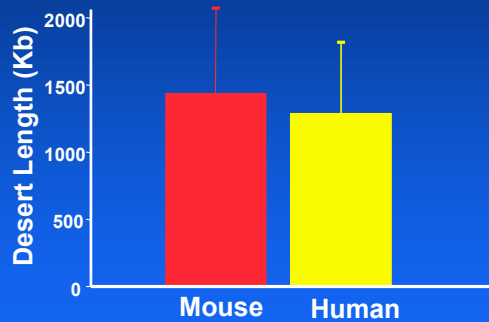
Orthologous Mouse “Gene Deserts”

- HUMAN
 - 234 Gene Deserts

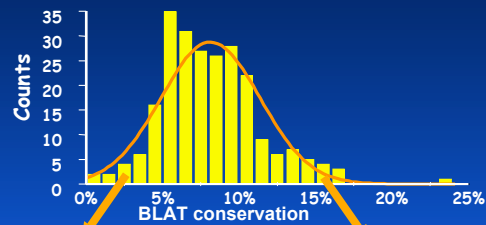
- Orthologous Mouse Comparison
 - 178 (74%) are also Deserts



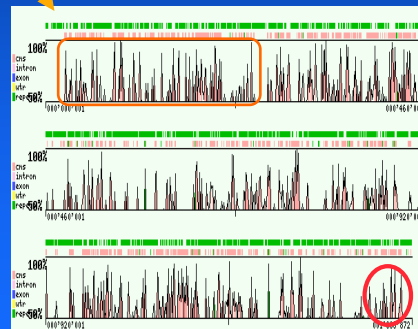
Human/Mouse “Gene Deserts” Comparisons



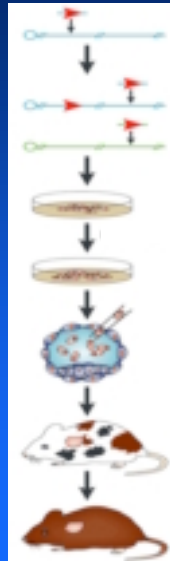
Human-Mouse Conservation in Gene Deserts



100 Kbp



Generation of Chromosomal Deletions in Mice



Insert loxP site into first end point

Insert loxP site into second end point
on same or different chromosome

Select and isolate double-targeted clones

Select and Isolate the ES-cell clones with the
desired rearrangements

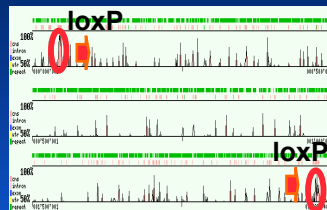
Inject ES cells into blastocyst

Identify chimeric mice

Derive progeny carrying
the engineered chromosome

Y. Yu and A. Bradley
Nat. Rev. Gen. 2, 780 -790 (2001)

Determining function of “Gene Deserts”



Cre-mediated deletion

